

Directed networks with underlying time structure from multivariate time series

Toshihiro Tanizawa,^{1,*} Tomomichi Nakamura,^{2,†} and Fumihiko Taya³

¹*Kochi National College of Technology, Monobe-Otsu 200-1, Nankoku, Kochi 783-8508, Japan*

²*Graduate School of Simulation Studies, University of Hyogo,*

7-1-28 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

³*Singapore Institute for Neurotechnology, Centre for Life Sciences,*

National University of Singapore, 28 Medical Drive, #05-COR, Singapore 117456

(Dated: May 11, 2015)

We discuss a versatile method of constructing directed networks from multivariate time series. While most common methods widely accepted at present utilize the concept of cross correlation between pairs of time series, the method presented here is based on the linear modeling technique in time series analysis. Since linear models generally contain terms representing feedback effects of different time delays, constructed networks reveal the intrinsic dynamical nature of the system under consideration such as complicated entanglement of different periodicities, which we referred to as “time structure”. The method enables us to construct networks even if a given multivariate time series do not have sufficiently large values of cross correlation, the case in which the approach using cross correlation is not applicable. We explicitly show a simple example where the method of cross correlation cannot reproduce the relationship among multivariate time series. The method we propose is demonstrated for numerical data generated by a known system and applied to two actual systems to see its effectiveness.

PACS numbers: 05.45.Tp, 89.20.Ff, 89.75.Fb, 89.75.Hc

Keywords: time series modeling, complex networks, directed networks, data visualization

I. INTRODUCTION

A time-dependent (dynamical) phenomenon produced by a complex system in the real world is, in most cases, an outcome of complicated interplay of many dynamical elements with different periodicities, time delays, or feedback effects. In this paper, we refer to the interplay as the “time structure” of the system. To gain a deeper insight into such a complicated phenomenon, a new frame of reference that enables us to view it from a different perspective often makes a significant contribution. It has been widely recognized that the concept of complex networks has a vast range of applicability and has given a new perspective on various problems [1–7]. In this paper we describe a generic method of constructing a directed network from a given set of multivariate time series generated by a (possibly nonlinear or stochastic) dynamical system based on a linear modeling technique including all relevant terms with different time delays. Hence, the networks are constructed from a dynamical system-wide perspective by the present method.

Regarding to the application of complex network theory to time series analysis, various algorithms for constructing networks from multivariate time series have been proposed [8–14]. In most of these algorithms, however, networks are constructed according to the values of the cross correlation between a pair of the time series [8–11, 13]. Although these methods have been proved to be effective for the applied problems, we consider that the

applicability is limited, because a collection of all pairwise correlations might not necessarily be the same as the many-body correlation of the system as a whole (the overall relationship), which reflects a complicated interplay among the dynamical elements with different periodicities, time delays, or feedback effects. In addition to this, the structure of the networks relies on arbitrary values of the threshold imposed on the cross correlation in making connections between nodes. Another approach by Gao and Jin [12] and Iwayama *et al.* [14] uses recurrence plots of time series. As this approach is also based on the collective information between a pair of time series data, it potentially contains the same deficiency as that of the method using pairwise cross correlation. The cross correlation is a useful statistic for investigating the degree of similarity between two data [15]. However, two data can have a relationship even if the cross correlation does not have large value (this is the case when two data are not similar). In this case, investigating the power spectrum is a useful alternative. Even when the dynamical behaviors of two data are different, if the power spectrums are the same, we can treat such data from the same population from the viewpoint of the linear theory [16, 17]. Several methods such as the method of Partial Directed Coherence (PDC) [18] and the method of Directed Transfer Function (DTF) [19] are based on this idea. In a sense, PDC and DTF methods treat multivariate time series as a whole. In estimating the power spectrums for a given set of multivariate time series, the vector auto-regressive (VAR) model is used, which is a simple extension of the AR (auto-regressive) model to multivariate cases. As we will discuss in Section III, however, the AR model has a problem of overfitting, which is also shared by the VAR model. Hence, the spectral resolution depends on the size

* tanizawa@ee.kochi-ct.ac.jp

† tomo@sim.u-hyogo.ac.jp

of VAR models. In addition, in PDC or DTF, nodes are connected when the power spectrums are similar, which is specified by an appropriately but arbitrarily chosen threshold. As a result, networks constructed by PDC and DTF methods are influenced by two things: (i) the limit in spectral resolution of the VAR model and (ii) an arbitrary value of the threshold. To avoid these difficulties, it might be preferable if we could find a simpler and more straightforward approach that enables us to capture many-body correlation among dynamical elements in a system as a whole.

The time series modeling is one of the well-established techniques for investigating various time-dependent phenomena. If a model successfully reproduces the behavior of given time series, it could be safe to say that the important features of the time structure built into the time series are appropriately taken into account by the model. Recently, Nakamura and Tanizawa proposed a method of constructing a weighted directed network based on a refined version of the auto-regressive (AR) model, the reduced auto-regressive (RAR) model [20–24], from a single (univariate) time series [25]. The network constructed by this method incorporates the entire time structure of the time series. On the edges of the network, positive real numbers (weights) are assigned based on the values of the coefficients of the linear model. Walker *et al.* partially applied this method to a system containing around 5000 variables for identifying critical components of a grain material controlling the collapse under strong shear strain [26]. Although multivariate time series are dealt with in the work of Walker *et al.*, the effects of different time delays in the time series are completely neglected in their network construction. One of the essential elements for understanding time-dependent phenomena is to incorporate the effects of underlying time delays or periodicities included in time series, which we refer to as “time structure” in this paper. Without including the effects of time structure, we cannot reproduce time-dependent phenomena [27]. However, to the best of the authors’ knowledge, an approach with a system-wide perspective including the time structure (dynamical system-wide perspective) has not yet been investigated properly.

The aim of this paper is therefore to present a general method for constructing directed networks from any given multivariate time series via time series modeling including all significant effects of time structure. Networks constructed by this method incorporate important effects of the time delays contained in the time series. As a method for multivariate time series modeling, we take the reduced auto-regressive (RAR) model, which is one of the well-established linear models [20–24]. Although the importance of taking nonlinearity into consideration is widely known, linear analyses still remain attractive and widely applied [19, 28, 29]. The key feature of the present approach is its combination of the linear model (in particular, the RAR model) containing terms with different time delays from multivariate time series and the directed network representation. While the two specific items by

themselves have been studied separately, the combination of these two enables us to unify the understanding of the dynamics and the structure of a complex system.

This paper is organized as follows. In the next section we discuss the shortcomings of the main ingredient of the common algorithms in network construction from time series, in which the relationship among a given set of time series is defined by the values of the pairwise cross correlation. In Section III, we explain our method by constructing a directed network based on the RAR model from multivariate time series generated by a given dynamical system. In Section IV, we apply our method to meteorological time series and electroencephalography (EEG) time series to see its effectiveness. We summarize the paper in Section V.

II. NETWORK CONSTRUCTION METHOD USING CROSS CORRELATION

In this section, we review the method for network construction according to the values of the cross correlation between pairs of multivariate time series. We also show a simple but general example in which this method cannot reproduce the original relationship among the dynamical elements.

A. Review of the common method

There are several works for constructing networks from multivariate time series [8–11, 13]. The common method upon which these works rely is the concept of the cross correlation. Generally, the basic procedure can be reduced to the following three steps.

1. Each time series is considered as a basic node of a network.
2. To investigate the relationship among multivariate time series, the cross correlation between each pair of time series (i.e. two time series) taken from the whole multivariate time series is calculated.
3. The pair of nodes corresponding to the chosen two time series is connected with an undirected edge when the value of the cross correlation is larger than an appropriately chosen threshold.

The large value of the cross correlation indicates that the pair of time series have some similarities or that both of these are under the influence of some common factors. It seems therefore logical to expect some sort of “relationship” between them. On the same footing, when two time series do not show any similarity, it also seems logical to conclude they are independent or have no relationship.

However, it should be emphasized that “similarity” is nonequivalent to “relationship”. For example, if two time series happen to be generated by two uncoupled identical oscillators, they would behave “similar”. On the

contrary, even if two time series show different time dependencies, it is still possible that they have some kind of complicated relationship. In addition, a relationship is, in many cases, directional. In other words, even if element α drives element β , it does not necessarily mean that element β drives back element α . This implies that the edges in the networks constructed from multivariate time series should have directions. The connections in the networks constructed using the cross correlation are undirected.

There is another problem. The cross correlation sometimes cannot precisely treat the total correlation among many elements. In other words, the collection of the cross correlations between all pairs of multivariate time series does not necessarily represent the entire correlation among the elements of the system. The patchwork of many two-body correlations might not be the same as the many-body correlation as a whole.

Therefore, even though the common method based on the concept of the cross correlation has been proved to be effective in various cases [8–11, 13], the range of applicability might be restrictive because of the above mentioned insufficiencies. Zalesky *et al.* have also pointed out that the cross correlation should be used cautiously in network construction [30]. In the next section, we show a simple but general example, where the method of the cross correlation cannot identify the explicit relationship among dynamical elements in a system.

B. A simple example revealing insufficiencies of the common method

Here we explicitly show a simple example that reveals fundamental insufficiencies of the common method. The system consists of four dynamical variables, $x_1(t)$, $x_2(t)$, $x_3(t)$ and $x_4(t)$, and their time dependencies are described by the following expressions:

$$x_1(t) = 0.7 x_1(t-1) - 0.4 x_1(t-3) + 0.3 x_2(t-4) + 0.2 x_4(t-7) + \varepsilon_1(t), \quad (1)$$

$$x_2(t) = 3.0 + 0.6 x_2(t-1) - 0.2 x_2(t-6) + \varepsilon_2(t), \quad (2)$$

$$x_3(t) = 0.5 x_1(t-2) + 0.3 x_4(t-9) + \varepsilon_3(t), \quad (3)$$

$$x_4(t) = 0.2 x_1(t-2) + 0.5 x_4(t-1) - 0.3 x_4(t-3) + \varepsilon_4(t), \quad (4)$$

where $\varepsilon_i(t)$ ($i = 1, 2, 3, 4$) are dynamic noise, independent and identically distributed Gaussian random variables with mean zero and standard deviation 1.0. The coefficients are chosen arbitrarily so that the generated time series do not diverge. In this paper, we distinguish “component” and “variable” as different technical terms. The term “component” is used for representing x_i and the term “variable” for representing $x_i(t-l)$ including its time delay. For instance, Eq. (1) has 3 components (x_1 , x_2 and x_4) and 4 variables, $x_1(t-1)$, $x_1(t-3)$, $x_2(t-4)$ and $x_4(t-7)$. As shown in Eqs. (1)–(4), each

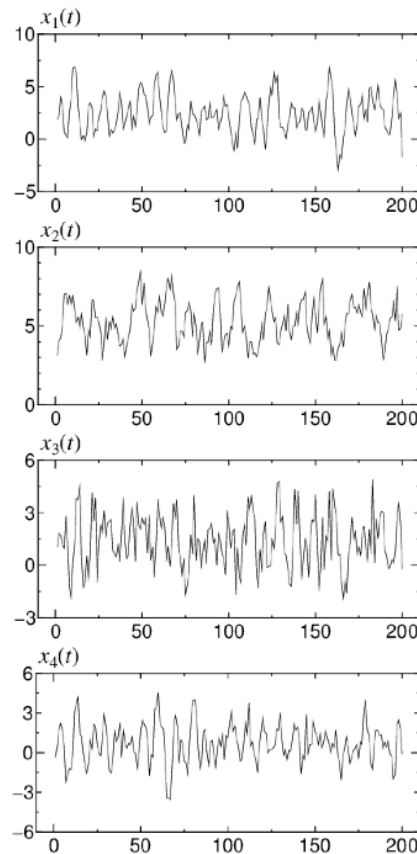


FIG. 1. Time series data generated by the linear models containing terms with different time delays defined by Eqs. (1)–(4). The first 200 data points for each component are plotted to show the time dependencies clearly.

dynamical variable at time t is determined by a linear combination of various other dynamical variables with discrete and different time delays. One realization of the four time series generated by these models are shown in Fig. 1.

Let us investigate the network constructed from these time series with the common method using the cross correlation. Since we have four time series corresponding to the components, x_1 , x_2 , x_3 , and x_4 , the network contains four nodes. We estimate the cross correlation (CC) of all pairs using 1000 data points generated by Eqs. (1)–(4) [31]. All the values are shown in Table I and the network constructed from these CCs with threshold 0.5 is shown in Fig. 2. With this threshold, only the nodes, x_1 and x_3 , are connected. The connection itself seems to be reasonable, because Eq. (3) representing the dynamics of $x_3(t)$ includes $x_1(t-2)$. However, Eq. (1) representing the dynamics of $x_1(t)$ does not include the component x_3 . The undirectedness of the connection thus cannot capture this one-way relationship between x_1 and x_3 . Furthermore, we would conclude that the pair, x_1 and x_4 are independent, since the value of CC between these components, 0.4452, is below the threshold 0.5. However, it is clearly untrue, because Eq. (4) representing

	x_1	x_2	x_3	x_4
x_1	1.0000	—	—	—
x_2	0.3608	1.0000	—	—
x_3	0.6692	0.1931	1.0000	—
x_4	0.4452	0.1790	0.4900	1.0000

TABLE I. The largest absolute values of the cross correlation of all possible pairs. The data are generated by Eqs. (1)–(4), and the values are estimated using 1000 data points.

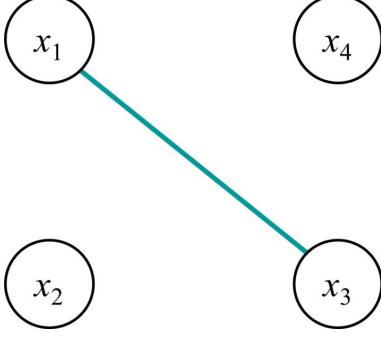


FIG. 2. The constructed network based on the values of the cross correlation shown in Table I with threshold 0.5. Contrary to the Eqs. (1)–(4), this network indicates that only x_1 and x_3 are connected and that x_2 and x_4 are independent.

the dynamics of $x_4(t)$ does include the variable $x_1(t-2)$. Although the connection between x_1 and x_4 is recovered by simply decreasing the value of the threshold from 0.5 to 0.4, this casual decrease seems to have no theoretical justification. This simple example shows two insufficiencies of the common method: (i) the undirectedness of the edges that cannot capture the directions of the relationship between components and (ii) the arbitrariness of the threshold value that cannot always recover existing relationships. In our opinion, the network constructed using the values of the cross correlation does not properly represent the exact relationship between components defined by Eqs. (1)–(4).

C. Preferable network structure

When we investigate a complicated time-dependent phenomenon generated by a vast number of elements, it is usually hopeless to obtain the exact expressions representing the entire time-dependency among them. In many cases, the only data available are multivariate time series. Despite of this fact, let us assume that we could obtain somehow such expressions as Eqs. (1)–(4) and investigate its straightforward network representation. Since the right hand side of Eq. (1), which determines the time dependency of the component x_1 , contains three components, x_1 , x_2 and x_4 , it seems logical to consider that x_1 is driven by these three components, x_1 , x_2 and x_4 . Similarly, from Eq. (2), x_2 is driven by only x_2 , from

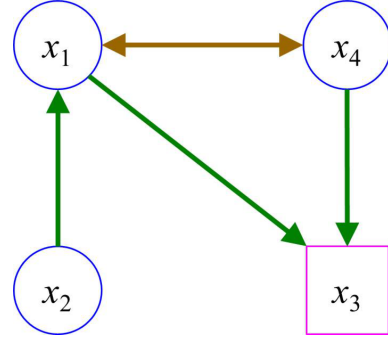


FIG. 3. (Color online) A preferable network structure according to the information of Eqs. (1)–(4). Notation \bigcirc means that the model for a component includes the component itself, and notation \square means that the component is not included in the model.

Eq. (3), x_3 is driven by x_1 and x_4 , and, from Eq. (4), x_4 is driven by x_1 and x_4 . A simple and direct representation of this relationship in the network construction would be a directed edge from node x_i to node x_j , when component x_i drives component x_j . The network structure constructed based on this idea is shown in Fig. 3. We consider that this structure appears to be a more faithful and straightforward network representation of the time structure of the system defined by Eqs. (1)–(4) than that constructed using the cross correlation with an arbitrary value of threshold. It should be noted that the network structure shown in Fig. 3 is free from the insufficiencies of the common method described in Section II B.

Once we somehow obtain a set of expressions describing the details of time dependencies among dynamical elements such as Eqs. (1)–(4), it is thus straightforward to find its network representation. To obtain such expressions, however, we must know or, at least, guess the microscopic mechanism that produces the observed time series, which is usually difficult and sometimes impossible. Since multivariate time series data are always obtained from any time-dependent phenomena by simple observation, it is crucial whether we can find dynamical expressions from observed time series without the knowledge of microscopic mechanism generating them. Fortunately, there are already well-established ways of obtaining such a set of linear expressions from any multivariate time series, one of which is the reduced auto-regressive (RAR) model [20]. The RAR model was originally devised as a refinement of linear modeling technique in time series analysis and has never been used as a base of network construction within our knowledge. In the next section we describe how to build an RAR model and construct a directed network from a given set of multivariate time series.

III. BUILDING A LINEAR MODEL FROM MULTIVARIATE TIME SERIES AND CONSTRUCTING A DIRECTED NETWORK

Periodic or nearly periodic behavior is an important nature for many time-dependent phenomena in the real world. Without including such (nearly) periodic effects, we cannot reproduce the time-dependent phenomena properly [27]. There are several widely accepted techniques to estimate the period of behavior, such as spectral estimation, auto-correlation, wavelet transforms and so on [32, 33]. All of these standard techniques either employ, are related to, or are a generalization of, Fourier series. Although the importance of taking nonlinearity into account is widely known in dealing with time series, linear analyses still remain attractive and widely applied [19, 28, 29].

One of the effective approaches to estimate periods built into complicated time series is to investigate the power spectrums using an auto-regressive (AR) model. The clear peaks found in the spectrums are identified as periods. The heights of the peaks are, however, depend on the frequency resolution of the AR model. If a dynamical phenomenon contains a large period, the AR model for describing the phenomenon needs to include terms with large time delay, which leads to the embedding in a high dimensional space, because AR models are built up from all terms with unit time difference up to the maximum time delay. Hence, the size of an AR model tends to become large. The AR modeling thus contains the problem of overfitting, although some terms contained in an AR model might not be necessary. Moreover, AR models contains all terms with unit time delay, we cannot know periodicities directly from the AR models.

Small and Judd have proposed a method to identify precise periodicities directly from the model [23]. The technique is based on an information theoretic reduction of AR models, which is referred to as the reduced auto-regressive (RAR) model [20–23]. RAR models include minimal number of terms indispensable for describing time series as assessed by an information criterion. Hence, RAR models can include terms with large time delay or terms acting as intermediate states like catalysts in chemical reactions, all of which might not be specified in AR models. Moreover, the RAR model has proven to be effective in modeling both linear and nonlinear dynamics [20–22]. There are also strong information theoretic arguments to support that the RAR model can detect any periodicities built into a given time series [23]. We therefore adopt here the RAR model for the basis of network construction.

The building of an RAR model from given time series proceeds as follows. Given a scalar time series $\{x(t)\}_{t=1}^n$ of n observations, an RAR model with the largest time

delay l_w can be expressed by

$$\begin{aligned} x(t) &= a_0 + a_1 x(t - l_1) + a_2 x(t - l_2) + \cdots + \\ &\quad a_w x(t - l_w) + \varepsilon(t) \\ &= a_0 + \sum_{i=1}^w a_i x(t - l_i) + \varepsilon(t), \end{aligned} \quad (5)$$

where $1 \leq l_1 < l_2 < \cdots < l_w$, a_i ($i = 0, 1, 2, \dots, w$) are parameters to be determined, and $\varepsilon(t)$ is assumed to be independent and identically distributed Gaussian random variables, which are interpreted as fitting errors. The parameters a_i are chosen to minimize the sum of the squares of fitting errors. To build an RAR model we prepare candidate basis functions used in the modeling, in the form of a dictionary, and select the most appropriate basis functions that can extract the peculiarities of the time series as much as possible. As RAR models are linear, the basis functions are a constant function and linear terms.

This methodology can be applied equally to multivariate time series straightforwardly [20–23]. A set of multivariate RAR models is expressed by

$$\begin{aligned} x_i(t) &= a_{i,0} + \sum_{j=1}^N \left(\sum_{k=1}^{w_j} a_{i,j,k} x_j(t - l_k) \right) + \varepsilon_i(t) \\ &\quad (i = 1, 2, \dots, N), \end{aligned} \quad (6)$$

where N is the number of components and $l_{w_i} (\geq 0)$ is the largest time delay of the i -th component. Initially, one constant term and $\sum_{j=1}^N w_j (\equiv L_{\text{init}})$ linear terms relating to time delay effects are prepared for building Eq. (6). We refer to these initial terms as the “dictionary of basis function” [20–23].

The RAR modeling approach has another essential procedure to extract necessary and sufficient terms from the initial $(1 + L_{\text{init}})$ terms for Eq. (6) by introducing the concept of description length in the information theory. The description length of a model of observed data consists of a term that captures how well the model fits the data and terms that penalize for the number of parameters needed for the model to achieve a prescribed accuracy. Specifically, an approximation to description length takes the form

$$\begin{aligned} DL(k) &= \left(\frac{n}{2} - 1 \right) \ln \frac{\mathbf{e}^T \mathbf{e}}{n} + \\ &\quad (k + 1) \left(\frac{1}{2} + \ln \gamma \right) - \sum_{i=1}^k \ln \delta_i \end{aligned} \quad (7)$$

where n is the length of the time series to be fitted, \mathbf{e} stands for the vector composed from fitting errors, k is the number of parameters (or model size), γ is related to the scale of the data, and the variables δ can be interpreted as the relative precision to which the parameters are specified. The factor γ is a constant and typically fixed to be $\gamma = 32$ [20]. From the initial dictionary containing $(1 + L_{\text{init}})$ terms, the most appropriate terms are

extracted to minimize this description length. More thorough arguments for the details of the RAR model can be found in [20, 21].

We note that selecting the optimal subset from a large dictionary of basis functions is an NP-hard problem that usually has to be solved heuristically [20]. For selecting basis functions, various algorithms have been proposed, which are proven to be effective in modeling both linear and nonlinear dynamics. The models obtained by these algorithms offers a near-optimal model [20, 21, 23, 24, 34]. In this paper, we use a selection algorithm using the total error [24, 34].

Let us apply the RAR modeling technique to the data represented in Fig. 1 to investigate whether we can reconstruct the system of Eqs. (1)–(4) [35]. In this case, we have four time series (that is, $x_1(t)$, $x_2(t)$, $x_3(t)$ and $x_4(t)$) of 1000 data points with Gaussian observational noise with the mean zero and the standard deviation 0.01. Choosing a time delay up to 10 for time series of each component and the constant function give 41 candidate basis functions in the initial dictionary [36]. Using the dictionary we build the multivariate RAR model for four components, x_1 , x_2 , x_3 , and x_4 . The result does indeed recover Eqs. (1)–(4), which verifies the consistency of the multivariate RAR model.

After building the multivariate RAR model corresponding to the systems under consideration, we use the information contained these models to construct a directed network representing the system. The model for the i -th variable $x_i(t)$ takes the form as

$$x_i(t) = a_{i,0} + a_{i,i,1}x_i(t-l_1) + a_{i,i,2}x_i(t-l_2) + a_{i,j,3}x_j(t-l_3) + a_{i,k,4}x_k(t-l_4) + e_i(t), \quad (8)$$

indicating that, to determine the value of x_i at time t , we need the information of the values of x_i , x_j , and x_k at some previous times. In our previous work on the network construction method from univariate time series [25], all information including the values of the parameters, $\{a_{i,j,w}\}$ and the time delays l_w were built into the network structure. In this paper regarding to the network construction method from multivariate times series, we focus on extracting the relationship among components to keep the argument simple and leave the treatment of the parameters and the time delays for future works. For this purpose, we pack the information of interdependency of the components contained in Eq. (8) into the form,

$$x_i = f_i(x_i, x_j, x_k), \quad (9)$$

representing that component x_i is a function of components, x_i , x_j and x_k , where f_i stands for the function representing the time dependency of the i -th component, x_i . When we construct a network from this expression, each component of the multivariate time series such as x_i is translated to a node. Next, we draw directed arrows from x_j to x_i and from x_k to x_i , if the right hand side of the model of x_i contains x_j and x_k . This basic idea enables us to construct a directed network embodying the

entire relationship among the components represented in a multivariate RAR model. In Eq. (9), the component x_i itself is included in the right hand side and the node x_i has a directed self-loop from x_i to x_i in the network. Such a case indicating that a component drives its own dynamics often happens. Since these self-loops make the network representation rather cumbersome, we represent the nodes with self-loops as circles and the node without self-loop as a square.

Let us apply the idea to the multivariate time series represented in Fig. 3, from which we obtain the multivariate RAR model exactly the same as Eqs. (1)–(4). The packed expressions such as Eq. (9) corresponding to this model become

$$x_1 = f_1(x_1, x_2, x_4), \quad (10)$$

$$x_2 = f_2(x_2), \quad (11)$$

$$x_3 = f_3(x_1, x_4), \quad (12)$$

$$x_4 = f_4(x_1, x_4). \quad (13)$$

Even these reduced expressions give us interesting information. Eqs. (10) and (11) tell that x_2 is autonomous and has influence on the dynamics of x_1 , although x_1 has no influence on the dynamics of x_2 . Eq. (12) tells that x_3 completely dependent on x_1 and x_4 , and does not drive its own dynamics at all. From Eqs. (10) and (13), we can understand that x_1 and x_4 are interconnected because both components are mutually included in the models.

Using this summarized information we construct a directed network, which is exactly the same as that shown in Fig. 3. Although it is true that the information contained in this network structure is no more than that contained in the packed expressions, Eqs. (10)–(13), the network representation becomes much more comprehensible as the number of the components increases. It should also be noted that it is impossible to obtain this network structure using the method with the cross correlation.

Network representation enables us to utilize various concepts in network theory such as the number of incoming edges (in-degree) and outgoing edges (out-degree) of each node. A large in-degree value of a node in the present representation means that the dynamics of the node is controlled by many other nodes. Similarly, a large out-degree value of a node means that the node has large influence on the dynamics of many other nodes. Let us evaluate the in-degree and out-degree of the directed network shown in Fig. 3. As node x_1 has two incoming arrows and the form of the node is \bigcirc as shown in Fig. 3 indicating a self-loop, the in-degree is 3. Also, as node x_1 has two outgoing arrows, the out-degree is 3 including the contribution from a self-loop. In the same manner, the in-degree and out-degree for other nodes x_2 , x_3 and x_4 are evaluated. Table II shows the numbers of in-degree and out-degree of all nodes.

As mentioned above, the method we propose is based on linear models built from time series. We understand that these models are phenomenological including parameters in the best predictive models and are not nec-

TABLE II. The number of in-degree and out-degree of the directed network shown in Fig. (3).

	x_1	x_2	x_3	x_4
in-degree	3	1	2	2
out-degree	3	2	0	3

essarily identical to the system that generates the dynamic phenomenon under consideration. In other words, the models might be mere approximation of the system. However, when the direct examination of the exact system is impossible and only limited information is available, even this approximated system gives us important clues to understand the phenomenon.

IV. APPLICATIONS

To show the wide range of applicability of the method described in the previous section, we present its application to two sets of multivariate time series taken from completely different research fields: (i) meteorological time series in Kobe, Japan and (ii) multichannel electroencephalography (EEG) time series with 10 channels. One of the reasons to consider these cases is that we can see the network structure and the relationship between nodes relatively easy because of their moderate system sizes. It should be emphasized, however, that the method can be applied to any larger systems straightforwardly.

A. Meteorological data in Kobe, Japan

Under controlled situations in laboratories, the relationship between atmospheric pressure and temperature is well described by the equation of state of gasses in thermodynamics. Actual situations in weather are, however, fairly complicated and far from controlled and the relationship is not obvious. With these circumstances in mind, we investigate the relationship using a set of meteorological data. The data we use here are five different time series: the atmospheric pressure, the atmospheric temperature, the dew-point temperature, the vapor pressure and the humidity, taken hourly in Kobe, Japan from from 1 January to 12 February in 2013 [37]. Figure 4 shows the profiles of the time series. Generally speaking, the relationship among these five time series is very complicated and hard to be extracted only from these data.

We use 1000 data points (around 42 days) to build multivariate RAR models. As there are 5 time series, choosing a time delay up to 15 for time series of each data and the constant function give 76 candidate basis functions in the dictionary. Using the dictionary we build the multivariate RAR model for each data. The reduced expressions of the obtained 5 multivariate RAR models

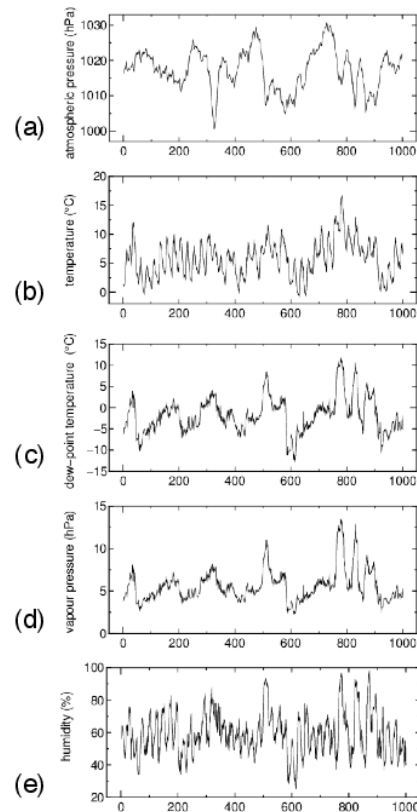


FIG. 4. Meteorological hourly time series in Kobe, Japan from 1 January to 12 February in 2013: (a) atmospheric pressure, (b) temperature, (c) dew-point temperature, (d) vapor pressure and (e) humidity. These data are used for building multivariate RAR models.

are

$$x_1 = f_1(x_1, x_2), \quad (14)$$

$$x_2 = f_2(x_2, x_4), \quad (15)$$

$$x_3 = f_3(x_3), \quad (16)$$

$$x_4 = f_4(x_3, x_4), \quad (17)$$

$$x_5 = f_5(x_2, x_3, x_5). \quad (18)$$

where x_1 corresponds to the atmospheric pressure, x_2 the atmospheric temperature, x_3 the dew-point temperature, x_4 the vapor pressure and x_5 the humidity.

In Fig. 5, we show the directed network constructed from these models representing the relationship of interdependency among these five data. Note that all models contain their own components, which means that all nodes in Fig. 5 have self-loops. The numbers of in-degree and out-degree for each node in Fig. 5 are shown in Table III. From these results we find that the atmospheric pressure and the humidity are influenced by others but do not have influence on any other components. On the contrary, the dew-point temperature is not influenced by others but has influence on other two components. The atmospheric temperature and the vapor pressure are influenced by others and have influence on others at the

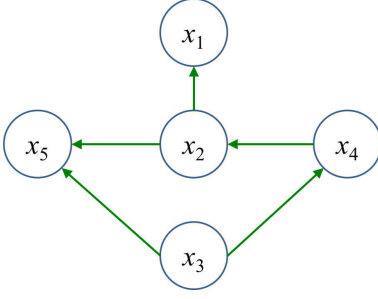


FIG. 5. (Color online) The directed network constructed by multivariate RAR models of meteorological data, where x_1 corresponds to atmospheric pressure, x_2 temperature, x_3 dew-point temperature, x_4 vapor pressure and x_5 humidity. All nodes are represented by \bigcirc , as all models contain their own components. For the explanation of the notation, see Fig. 3.

TABLE III. The number of in-degree and out-degree of the directed network of meteorological data shown in Fig. (5).

	atmospheric pressure	temp.	dew-point temp.	vapor pressure	humidity
in-deg.	1	1	0	1	2
out-deg.	0	2	2	1	0

same time. We also found that there is no mutual (bi-directional) connections between any component.

B. Electroencephalogram (EEG) data

The second application is to construct network for electroencephalogram (EEG) data [38]. The mammalian brain is certainly one of the most complex systems in nature. It is made of billions of cells endowed with individual electrical activity and interconnected in a highly intricate network. The average electrical activity of a portion of this network may be recorded in the course of time and is called the electroencephalogram (EEG) [39]. It is of immense value to understand the activity of the human brain using EEG data [40]. The EEG signal we use here was recorded from a healthy human adult under a condition (eyes closed and resting) in a shield room. The EEG data were simultaneously obtained from 10 channels of Fz, Cz, Pz, Oz, F3, F4, C3, C4, P3 and P4 of the unipolar 10-20 Jasper registration scheme [41]. See Fig. 6 for the placement of the electrodes. Here, the capital letters, F, C, P, and O, stand for the frontal, central, parietal, and occipital lobes in the brain, respectively. The data were digitized at 1024[Hz] using a 12-bit A/D converter. The EEG impedances were less than 5[K Ω]. The data were amplified by gain = 18 000, and amplifier frequency cut-off settings of 0.03[Hz] and 200[Hz] were used [42]. We use 1000 data points (around 1 second) to build multivariate RAR models. Figure 7 shows time

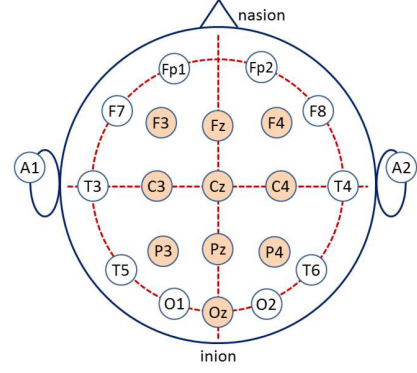


FIG. 6. (Color online) The International 10-20 system of electrode placement. We use data from 10 channels, Fz, Cz, Pz, Oz, F3, F4, C3, C4, P3 and P4.

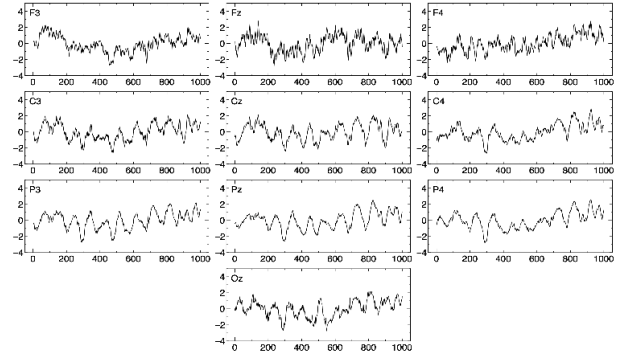


FIG. 7. Multichannel electroencephalography (EEG) time series of Fz, Cz, Pz, Oz, F3, F4, C3, C4, P3 and P4 used for building multivariate RAR models.

series of each channel.

As there are 10 channels, choosing a time delay up to 25 for time series of each channel and the constant function give 251 candidate basis functions in the dictionary. Using the dictionary we build the multivariate RAR model for each channel. The reduced expressions of the obtained 10 multivariate RAR models are

$$Fz = f_{Fz}(Fz, Cz), \quad (19)$$

$$Cz = f_{Cz}(Cz, Pz, Oz, F3), \quad (20)$$

$$Pz = f_{Pz}(Fz, Cz, Pz, Oz, C3, P3, P4), \quad (21)$$

$$Oz = f_{Oz}(Fz, Cz, Oz), \quad (22)$$

$$F3 = f_{F3}(F3), \quad (23)$$

$$F4 = f_{F4}(Fz, F4), \quad (24)$$

$$C3 = f_{C3}(F3, C3, P3), \quad (25)$$

$$C4 = f_{C4}(Fz, C4, P4), \quad (26)$$

$$P3 = f_{P3}(Pz, Oz, P3), \quad (27)$$

$$P4 = f_{P4}(Fz, Cz, Pz, P4). \quad (28)$$

In Fig. 8, we show the directed network constructed from these models representing the relationship of interdependency among these 10 channels. Note that all

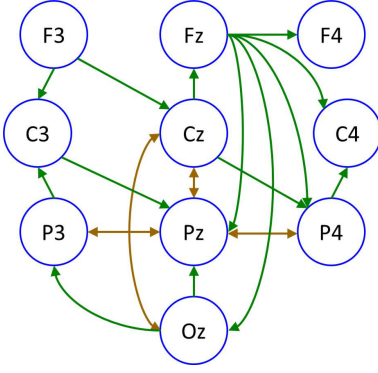


FIG. 8. (Color online) The directed network constructed by multivariate RAR models from Eq. (19) to Eq. (28). All nodes are represented by \bigcirc , as all models contain their own components. For the explanation of the notation, see Fig. 3.

TABLE IV. The numbers of in-degree and out-degree for each node of the directed network of EEG data shown in Fig. 8.

	Fz	Cz	Pz	Oz	F3	F4	C3	C4	P3	P4
in-degree	2	4	7	3	1	2	3	3	3	4
out-degree	6	5	4	4	3	1	2	1	3	3

models contain their own components, which means that all nodes in Fig. 8 have self-loops. In Fig. 8, the channels, Cz, Pz, P3, P4, and Oz, are mutually connected by bi-directional arrows.

The numbers of in-degree and out-degree for each node in Fig. 8 are shown in Table IV. In Fig. 9, we show the directed networks, the same as that showed in Fig. 8, with each node size is proportional to the numbers of incoming arrows (Fig. 9(a)) and outgoing arrows (Fig. 9(b)). In Fig. 9(a), the size of node Pz is the largest. Since an incoming arrow into a node implies that the dynamics of the node is partially governed by the source node of the arrow, the dynamics of node Pz that belongs to the parietal lobe is determined under the influences of many other parts of the brain. Since node Pz is the part that integrates sensory information from all parts of the brain, this result seems admissible. On the other hand, the size of node Fz is the largest in Fig. 9(b). Since an outgoing arrow from a node implies that the node partially controls the dynamics of the target node, it implies that node Fz belonging to the frontal lobe is the most influential on the dynamics of the other nodes in the brain. It is striking that we can extract such an amount of information about the brain functioning without any physiological knowledge only from the combination of the RAR model and the network visualization scheme.

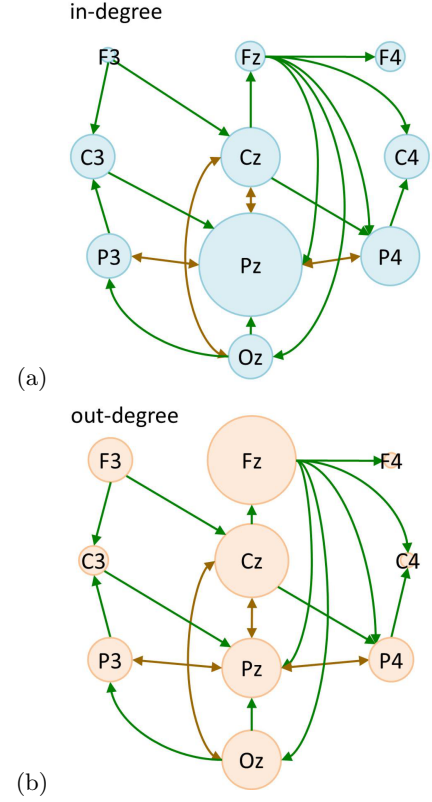


FIG. 9. (Color online) Directed networks represented in Fig. 8 with the adjusted node size: (a) the node size is proportional to the number of in-degree and (b) to the number of out-degree. An outgoing arrow from a node implies that the node partially controls the dynamics of the target node and an incoming arrow into a node implies that the dynamics of the node is partially governed by the source node of the arrow.

V. SUMMARY

In summary, we have described a generic algorithm for constructing directed networks from multivariate time series based on the RAR modeling technique. The strong point of this method is that it enables us to extract the hidden relationship among dynamical components from a dynamical system-wide perspective even if the time series do not have large values of cross correlation. The edges in the networks constructed by the method have directions representing the dependencies of the relationship of controlling and controlled. It should also be noted that the method can be applicable to any kind of multivariate time series without knowledge of underlying relationship among the components of the dynamical system under consideration in advance. The relationship can be extracted after the construction of the network. We have applied the method to meteorological time series and multichannel electroencephalography (EEG) data. The networks constructed in the application reveal the hidden relationships among components in the data.

VI. ACKNOWLEDGMENTS

The authors thank to the anonymous referees for various valuable comments and suggestions. Toshi Tanizawa would like to acknowledge the support of a Grant-in-Aid for Scientific Research (C) (No. 24540419) from the

Japan Society for the Promotion of Science (JSPS). Tomo Nakamura and the other authors would like to acknowledge Professor Paul E. Rapp (Uniformed Services University of the Health Sciences) for providing us with the EEG data.

-
- [1] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
 - [2] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
 - [3] R. Albert and A.-L. Barabási, *Reviews of Modern Physics* **74**, 47 (2002).
 - [4] M. E. J. Newman, *SIAM Review* **45**, 167.
 - [5] J. Zhang and M. Small, *Physical Review Letters* **96**, 238701 (2006).
 - [6] M. Small, D. M. Walker, and C. K. Tse, *Physical Review Letters* **99**, 188702 (2007).
 - [7] M. Small, J. Zhang, and X. Xu, in *Complex Sciences* (Springer Berlin Heidelberg, Shanghai, China, 2009) pp. 2078–2089.
 - [8] R. N. Mantegna, *The European Physical Journal B* **11**, 193 (1999).
 - [9] I. Farkas, H. Jeong, T. Vicsek, A.-L. Barabási, and Z. N. Oltvai, *Physica A* **318**, 601 (2003).
 - [10] K. Yamasaki, A. Gozolchiani, and S. Havlin, *Physical Review Letters* **100**, 228501 (2008).
 - [11] A. Tsonis and K. Swanson, *Physical Review Letters* **100**, 228502 (2008).
 - [12] Z. Gao and N. Jin, *Chaos* **19**, 033137 (2009).
 - [13] C. K. Tse, J. Liu, and F. C. M. Lau, *Journal of Empirical Finance* **17**, 659 (2010).
 - [14] K. Iwayama, Y. Hirata, K. Takahashi, K. Watanabe, K. Aihara, and H. Suzuki, *Scientific Reports* **2**, 423 (2012).
 - [15] Broadly speaking, we use the term “similarity” when the time evolution or power spectrum patterns of two dynamical elements are almost identical.
 - [16] A. Galka, *Topics in Nonlinear Time Series Analysis* (World Scientific Publishing Company, Singapore, 2000).
 - [17] T. Schreiber and A. Schmitz, *Physica D* **142**, 346 (2000).
 - [18] L. A. Baccalá and K. Sameshima, *Biological Cybernetics* **84**, 463 (2001).
 - [19] L. Astolfi, F. Cincotti, D. Mattia, M. G. Marciani, L. A. Baccalá, F. de Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. C. Edgar, G. A. Miller, B. He, and F. Babiloni, *Human brain mapping* **28**, 143 (2007).
 - [20] K. Judd and A. Mees, *Physica D* **82**, 426 (1995).
 - [21] K. Judd and A. Mees, *Physica D* **120**, 273 (1998).
 - [22] M. Small and K. Judd, *Physica D* **117**, 283 (1998).
 - [23] M. Small and K. Judd, *Physical Review E* **59**, 1379 (1999).
 - [24] T. Nakamura, D. Kilminster, K. Judd, and A. Mees, *International Journal of Bifurcation and Chaos* **14**, 1129 (2004).
 - [25] T. Nakamura and T. Tanizawa, *Physica A* **391**, 4704 (2012).
 - [26] D. M. Walker, A. Tordesillas, T. Nakamura, and T. Tanizawa, *Physical Review E* **87**, 032203 (2013).
 - [27] T. Ohira and T. Yamane, *Physical Review E* **61**, 1247 (2000).
 - [28] X. Fu, M. Small, D. M. W. Walker, and H. Zhang, *Physical Review E* **77**, 036113 (2008).
 - [29] C. Rummel, E. Abela, M. Müller, M. Hauf, O. Scheidegger, R. Wiest, and K. Schindler, *Physical Review E* **83**, 066215 (2011).
 - [30] A. Zalesky, A. Fornito, and E. Bullmore, *NeuroImage* **60**, 2096 (2012).
 - [31] G. Kitagawa, *Introduction to Time Series Modeling* (Chapman and Hall/CRC, New York, 2010).
 - [32] M. B. Priestley, *Non-linear and Non-stationary Time Series Analysis* (Academic Press, London, 1988).
 - [33] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer* (Prentice Hall, Upper Saddle River, N.J., 1998).
 - [34] T. Nakamura and M. Small, *Physica D* **223**, 54 (2006).
 - [35] We first apply the bottom-up method using the total error [24]. The best model might be trapped in one of the local minima. To avoid this situation and to reduce the problem to a manageable size, a model is built from the complete dictionary using the top-down method, but starting from the model whose size is 10 larger than that of the best model obtained by the bottom-up method. We select the model as the best model whose description length is the smallest. It should be noted that this method do no worse than the bottom-up method.
 - [36] These are the constant function and $x_1(t-1)$, $x_1(t-2)$, \dots , $x_1(t-10)$, $x_2(t-1)$, \dots , $x_2(t-10)$, $x_3(t-1)$, \dots , $x_3(t-10)$, $x_4(t-1)$, \dots , $x_4(t-10)$.
 - [37] The data can be obtained from Japan Meteorological Agency, <http://www.jma.go.jp/jma/indexe.html>.
 - [38] K. Lehnertz, G. Ansmann, S. Bialonski, H. Dickten, C. Geier, and S. Porz, *Physica D* **267**, 7 (2014).
 - [39] W. J. Freeman, *Mass Action in the Nervous System* (Academic Press, New York, 1975).
 - [40] We understand that EEG time series have already been analyzed with various techniques including network visualization. Regarding to the research of the brain from the perspective of network representation, see for example, a concise review article by Lehnertz *et al.* [38] and the references therein. The purpose for presenting the application here is not to claim the superiority of our method over other methods in studying EEG but to show the versatility of our method. Our aim is not to overinterpret the results of EEG in this paper.
 - [41] H. H. Jasper, *Electroencephalography and Clinical Neurophysiology* **10**, 371 (1958).
 - [42] T. A. A. Watanabe, C. J. Cellucci, E. Kohegyi, T. R. Bashore, R. C. Josiassen, N. N. Greenbaun, and P. E. Rapp, *Psychophysiology* **40**, 77 (2003).